

RESEARCH ARTICLE

# Genetic determinants of gut microbiota composition and bile acid profiles in mice

Julia H. Kemis<sup>1</sup>, Vanessa Linke<sup>2</sup>, Kelsey L. Barrett<sup>1</sup>, Frederick J. Boehm<sup>3</sup>, Lindsay L. Traeger<sup>1</sup>, Mark P. Keller<sup>4</sup>, Mary E. Rabaglia<sup>4</sup>, Kathryn L. Schueler<sup>4</sup>, Donald S. Stapleton<sup>4</sup>, Daniel M. Gatti<sup>5</sup>, Gary A. Churchill<sup>5</sup>, Daniel Amador-Noguez<sup>1</sup>, Jason D. Russell<sup>2</sup>, Brian S. Yandell<sup>3</sup>, Karl W. Broman<sup>6</sup>, Joshua J. Coon<sup>2,7,8</sup>, Alan D. Attie<sup>4</sup>, Federico E. Rey<sup>1\*</sup>

**1** Department of Bacteriology, University of Wisconsin–Madison, Madison, Wisconsin, United States of America, **2** Department of Chemistry, University of Wisconsin–Madison, Madison, Wisconsin, United States of America, **3** Department of Statistics, University of Wisconsin–Madison, Madison, Wisconsin, United States of America, **4** Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin, United States of America, **5** Jackson Laboratory, Bar Harbor, Maine, United States of America, **6** Department of Biostatistics and Medical Informatics, University of Wisconsin–Madison, Madison, Wisconsin, United States of America, **7** Morgridge Institute for Research, Madison, Wisconsin, United States of America, **8** Department of Biomolecular Chemistry, University of Wisconsin–Madison, Madison, Wisconsin, United States of America

\* [ferey@wisc.edu](mailto:ferey@wisc.edu)



**OPEN ACCESS**

**Citation:** Kemis JH, Linke V, Barrett KL, Boehm FJ, Traeger LL, Keller MP, et al. (2019) Genetic determinants of gut microbiota composition and bile acid profiles in mice. *PLoS Genet* 15(8): e1008073. <https://doi.org/10.1371/journal.pgen.1008073>

**Editor:** Cisca Wijmenga, UMC Groningen, NETHERLANDS

**Received:** March 4, 2019

**Accepted:** June 14, 2019

**Published:** August 29, 2019

**Copyright:** © 2019 Kemis et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data reported in this paper are accessible in the NCBI Short Read Archive (SRP) under accession ID PRJNA492330. Mass spectrometry data files are available on Chorus ([chorusproject.org](https://chorusproject.org)) accession number: project ID 1568.

**Funding:** This work was supported by the National Institutes of Health (NIH) grants DK108259 (F.E.R.), DK101573 (A.D.A.), GM070683 (K.W.B. and G.A.C.), NIH National Institute of Allergy and Infectious Diseases grant T32AI55397 (J.H.K.),

## Abstract

The microbial communities that inhabit the distal gut of humans and other mammals exhibit large inter-individual variation. While host genetics is a known factor that influences gut microbiota composition, the mechanisms underlying this variation remain largely unknown. Bile acids (BAs) are hormones that are produced by the host and chemically modified by gut bacteria. BAs serve as environmental cues and nutrients to microbes, but they can also have antibacterial effects. We hypothesized that host genetic variation in BA metabolism and homeostasis influence gut microbiota composition. To address this, we used the Diversity Outbred (DO) stock, a population of genetically distinct mice derived from eight founder strains. We characterized the fecal microbiota composition and plasma and cecal BA profiles from 400 DO mice maintained on a high-fat high-sucrose diet for ~22 weeks. Using quantitative trait locus (QTL) analysis, we identified several genomic regions associated with variations in both bacterial and BA profiles. Notably, we found overlapping QTL for *Turicibacter sp.* and plasma cholic acid, which mapped to a locus containing the gene for the ileal bile acid transporter, *Slc10a2*. Mediation analysis and subsequent follow-up validation experiments suggest that differences in *Slc10a2* gene expression associated with the different strains influences levels of both traits and revealed novel interactions between *Turicibacter* and BAs. This work illustrates how systems genetics can be utilized to generate testable hypotheses and provide insight into host-microbe interactions.

NLM Computation and Informatics in Biology and Medicine Postdoctoral Fellowship 5T15LM007359 (L.L.T.), NIH National Institute of General Medical Sciences T32GM008349 (K.L.B.), and Transatlantic Networks of Excellence Award from the Leducq Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Author summary

Inter-individual variation in the composition of the intestinal microbiota can in part be attributed to host genetics. However, the specific genes and genetic variants underlying differences in the microbiota remain largely unknown. To address this, we profiled the fecal microbiota composition of 400 genetically distinct mice, for which genotypic data is available. We identified many loci of the mouse genome associated with changes in abundance of bacterial taxa. One of these loci is also associated with changes in the abundance of plasma bile acids—metabolites generated by the host that influence both microbiota composition and host physiology. Follow up validation experiments provide mechanistic insights linking host genetic differences, with changes in ileum gene expression, bile acid-bacteria interactions and bile acid homeostasis. Together, this work demonstrates how genetic approaches can be used to generate testable hypothesis to yield novel insight into how host genetics shape gut microbiota composition.

## Introduction

The intestinal microbiota has profound effects on host physiology and health [1–3]. The composition of the gut microbiota is governed by a combination of environmental factors, including diet, drugs, maternal seeding, cohabitation, and host genetics [4–7]. Together, these factors cause substantial inter-individual variation in microbiota composition and modulate disease risk [8,9]. Alterations in the composition of the microbiota are associated with a spectrum of cognitive, inflammatory and metabolic disorders [10–12], and a number of bacterial taxa have been causally linked with modulation of disease [13–15]. A major challenge in the field is deciphering how host genetics and environmental factors interact to shape the composition of the gut microbiota. This knowledge is key for designing strategies aimed at modifying gut microbiota composition to improve health outcomes.

Several mouse and human studies have examined the role of host genetics in shaping the composition of the gut microbiota [16]. Mouse studies comparing gut bacterial communities from inbred mouse strains [17,18] and strains harboring mutations in immune-related genes [19–22] support this notion. Additionally, quantitative trait locus (QTL) analyses in mice have identified genetic regions associated with the abundance of several bacterial taxa and community structure [23–26]. Twin studies and genome-wide association studies (GWAS) in humans have identified heritable bacterial taxa and SNPs associated with specific gut microbes. While comparing these studies is often difficult due to differences in environmental variables among populations, some associations are consistently detected among geographically discrete populations, such as the association between *Bifidobacterium* abundance and the lactase (*LCT*) gene locus [27–29], indicating the abundance of specific taxa is influenced by host genetic variation.

Gut microbes and the host communicate through the production and modification of metabolites, many of which impact host physiology [30–34]. Bile Acids (BAs) are host-derived and microbial-modified metabolites that regulate both the gut microbiome and host metabolism [35–37]. BAs are synthesized in the liver from cholesterol, stored in the gallbladder and are secreted in the proximal small intestine where they facilitate absorption of fat-soluble vitamins and lipids. Once in the intestine, BAs can be metabolized by gut bacteria through different reactions, including deconjugation, dehydroxylation, epimerization, and dehydrogenation, to produce secondary BAs with differential effects on the host [33,35]. In addition to their direct effects on the host, BAs shape the gut microbiota composition through antimicrobial activities [38,39]. The detergent properties of BAs cause plasma membrane damage. The

bactericidal activity of a BA molecule corresponds to its hydrophobicity [40]. Additionally, the microbiota modulates primary BA synthesis through regulation of the nuclear factor FXR [41]. Thus, we hypothesized that host genetic variation associated with changes in BA homeostasis mediates alterations in gut microbiota composition.

To investigate how genetic variation affects gut microbiota and BA profiles, we used the Diversity Outbred (DO) mouse population, which is a heterogeneous population derived from eight founder strains: C57BL6/J (B6), A/J (A/J), 1291/SvImJ (129), NOD/ShiLtJ (NOD), NZO/HiLtJ (NZO), CAST/EiJ (CAST), PWK/PhJ (PWK), and WSB/EiJ (WSB) [42,43]. These eight strains capture a large breadth of the genetic diversity found in inbred mouse strains. Additionally, the founder strains harbor distinct gut microbial communities and exhibit disparate metabolic responses to diet-induced metabolic disease [18,44,45]. The DO population is maintained by an outbreeding strategy aimed at maximizing the heterozygosity of the outbred stock. The genetic diversity and large number of generations of outbreeding make it an ideal resource for high-resolution genetic mapping of microbial and metabolic traits [43].

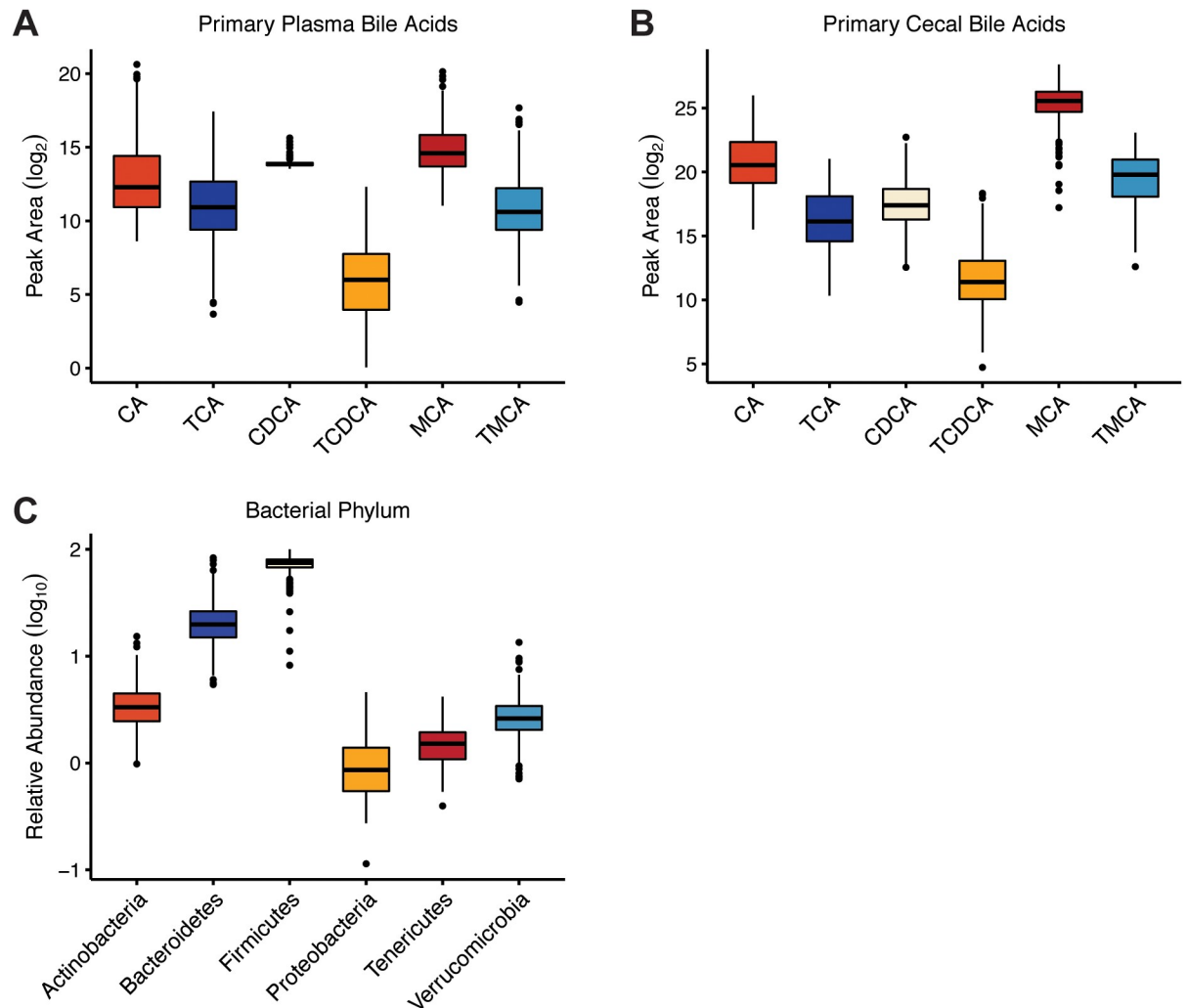
We characterized the intestinal microbiota composition and plasma and cecal BA profiles in ~400 genetically distinct DO mice fed a high-fat/high-sucrose diet for ~22 weeks and performed quantitative trait loci (QTL) analysis to identify host genetic loci associated with these traits. Specifically, we focused our analysis on potentially pleiotropic loci, which we defined as a single genetic locus that associates with both bacterial and BA traits. Our analysis revealed several instances of bacterial and metabolite traits attributed to the same DO founder haplotypes mapping to the same position of the mouse genome, including a locus associated with plasma BA levels and the disease-modulating organism *Akkermansia muciniphila*. Additionally, we identified the ileal BA transporter *Slc10a2* as a candidate gene that regulates both the abundance of *Turicibacter sp.* and plasma levels of cholic acid.

## Results and discussion

### Phenotypic variation among Diversity Outbred (DO) mice fed high-fat and high-sucrose diet

We investigated the impact of genetic variation on gut microbiota composition and bile acid (BA) profiles using a cohort of ~400 DO mice maintained on a high-fat high-sucrose diet (45% kcal from fat and 34% from sucrose) for ~22 weeks (range 21–25 weeks), starting at weaning. We previously showed that this diet elicits a wide range of metabolic responses in the eight founder strains that are associated with microbiome changes [18,46]. Furthermore, we incorporated in our analyses previously published clinical weight traits collected from the same DO mice [47]. All animals were individually housed throughout the duration of the study to measure food intake and minimize microbial exchange.

We performed LC-MS/MS analyses of plasma and cecal contents to assess relative variation in the levels of 27 BAs. Both plasma and cecal bile acids were measured to provide a comprehensive picture of systemic BA homeostasis. There was substantial variation in the plasma and cecal BA profiles across the 384 mice (Fig 1A and 1B; S1 Table). Additionally, we examined gut microbiota composition (n = 399) using 16S rRNA gene amplicon sequencing of DNA extracted from fecal samples collected at the end of the experiment. Within the cohort, there were 907 unique Exact Sequence Variants (ESVs), (100% operational taxonomic units defined with dada2 [48]), which were agglomerated into 151 lower taxonomic rankings (genus, family, order, class, phyla). The microbial traits represented each of the major phyla found in the intestine and the relative abundance of these phyla was highly variable among the DO mice (Fig 1C). For instance, the abundance of taxa classified to the Bacteroidetes phylum ranged from 1.17–89.28%.



**Fig 1. Phenotypic variation among Diversity Outbred (DO) mice fed high-fat and high-sucrose diet.** (A) Abundance (peak area) of primary bile acids detected in plasma and (B) cecal contents (n = 384). (C) Distributions of the normalized relative abundance of bacterial phyla identified in DO fecal microbiota (n = 399).

<https://doi.org/10.1371/journal.pgen.1008073.g001>

For subsequent analysis, we identified a core measurable microbiota (CMM), which we defined as taxon found in at least 20% of the mice [24]. This was done to remove the effects of excessive variation in the data due to bacterial taxa that were low abundance and/or sparsely distributed. In total, the CMM was comprised of 86 ESVs and 42 agglomerated taxa (S2 Table). The CMM traits represent a small fraction of the total microbes detected, but account for 94.5% of the rarefied sequence reads, and therefore constitute a significant portion of the identifiable microbiota.

Since mice were received in cohorts (i.e., waves) of 100, we examined whether animals in each wave were more similar to each other than mice in other waves. The fecal microbiota composition significantly clustered by wave ( $p < 0.001$ , PERMANOVA) and sex ( $p < 0.001$ , PERMANOVA) (S1 Fig). PCA analysis of plasma and cecal bile acids showed a significant effect of sex, but not wave, on both plasma ( $p < 0.0001$ , Kruskal Wallis) and cecal BA profiles ( $p < 0.05$ , Kruskal Wallis) (S2 Fig).

There is substantial evidence implicating gut microbiota and BAs in metabolic disease development [36,37]. To identify potential relationships among these traits, we performed

correlation analysis which yielded many significant associations after FDR correction (FDR < 0.05) (S3 Table, discussed in S1 Data).

## Abundance of gut bacterial taxa and bile acids are associated with host genetics

To identify associations between regions of the mouse genome and the clinical and molecular traits discussed above, we performed QTL analysis using the R/qtl2 package [49]. We used sex, days on the diet, and experimental wave as covariates. We identified 13 significant QTL (LOD  $\geq$  7.66;  $P \leq$  0.05) and 50 suggestive QTL (LOD  $\geq$  6.80;  $P \leq$  0.2) for bacterial [36], bile acid [13], and body weight [1] traits (Fig 2, S4 Table).

Of the microbial QTL, we found 23 QTL for 17 distinct bacterial ESVs from the Bacteroidetes and Firmicutes phyla that met the LOD  $\geq$  6.80 threshold. ESVs with the strongest QTL (LOD > 8) are classified to the Clostridiales order and map on chr 12 at ~33 Mbp, the Lachnospiraceae family on chr 2 at 164 Mbp, and the S24-7 family on chr 2 at ~115 Mbp. We also identified 12 QTL for microbial taxa collapsed by taxonomic assignment (i.e., genus to phylum). The genera *Lactococcus* and *Oscillospira* were also associated with host genetic variation, which is consistent with previous studies [23,24,50,51].

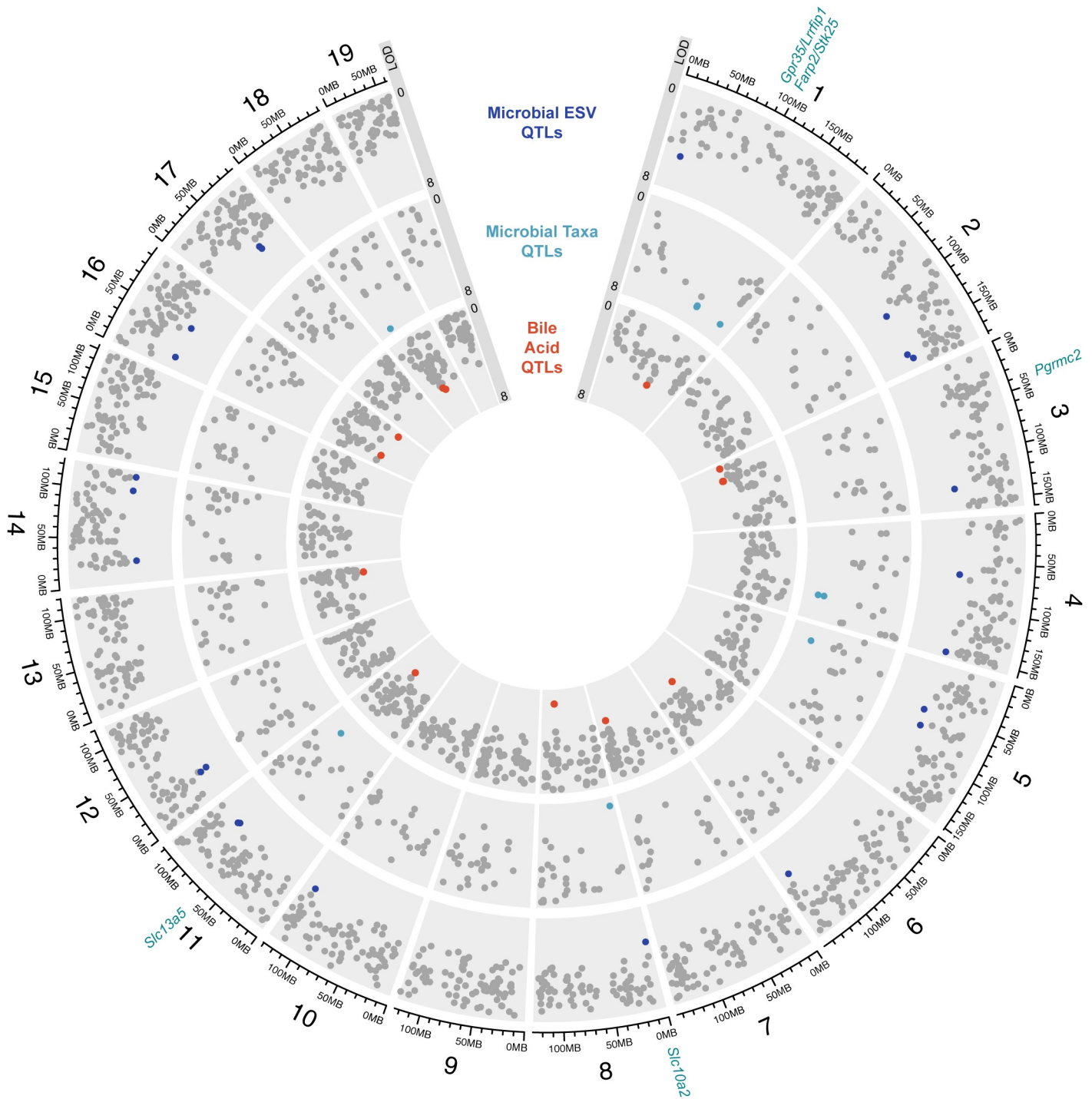
Similarly, BA QTL mapped to multiple loci spanning the mouse genome and most BA traits mapped to multiple positions. BA synthesis and metabolism are regulated by multiple host signaling pathways: there are >17 known host enzymes involved in the production of BAs [36], transporters, which play a critical role in maintaining the enterohepatic circulation and BA homeostasis, and receptors that respond to BA in a variety of host tissues [52–54]. Therefore, it is not surprising that our results indicate that BA levels are polygenic and shaped by multiple host factors.

To identify instances of overlapping QTL, we applied a less stringent threshold of LOD  $\geq$  6.1 ( $P <$  0.5). We observed multiple instances of related BA species associating to the same genetic locus, indicating the presence of pleiotropic loci. Interestingly, several of these loci associate with levels of related BA species in different stages of microbial modification. For example, cecal taurocholic acid (TCA) and plasma CA QTL overlap on chr 7 at 122 Mbp. Likewise, QTL for plasma TDCA and cecal DCA, overlap on chr 12 between ~99–104 Mbp. For the cecal DCA, the WSB founder haplotype was associated with higher levels of this BA, while the NOD founder haplotype was associated with lower levels. The opposite pattern was observed for plasma TDCA, where the NOD and WSB haplotype were associated with higher and lower levels, respectively (S3A and S3B Fig).

We also identified overlapping QTLs on chr 11 at ~71 Mbp for cecal levels of the secondary BAs lithocholic acid (LCA) and isolithocholic acid (ILCA), the isomer of LCA produced by bacterial epimerization (S3C Fig). Higher levels of these cecal BAs are associated with the 129 founder haplotype and lower levels are associated with the A/J founder haplotype (S3D and S3E Fig). We identified the positional candidate gene *Slc13a5* (S3F Fig), which is a sodium-dependent transporter that mediates cellular uptake of citrate, an important precursor in the biosynthesis of fatty acids and cholesterol [55]. Recent evidence indicates that *Slc13a5* influences host metabolism and energy homeostasis [56–58]. *Slc13a5* is a transcriptional target of pregnane X receptor (PXR) [59], which also regulates the expression of genes involved in the biosynthesis, transport, and metabolism of BAs [60].

## Co-mapping analyses identify novel interactions between bacterial taxa and bile acid homeostasis

We searched for regions of the chromosome that were associated with both BA and bacterial abundance, as this may provide evidence of interactions between the traits [61]. We identified



**Fig 2. Genetic architecture of quantitative trait loci (QTL) for microbial exact sequence variants (ESVs) and taxa abundance, and plasma and cecal bile acids in 400 Diversity Outbred (DO) mice.** The outer layer shows the chromosome location where major tick marks correspond to 25 Mbp. Logarithm of the odds (LOD) range is shown for each track. Each dot represents a QTL on each chromosome of the mouse genome for a given trait. Grey dots denote QTLs with LOD < 6.8. Candidate genes discussed in text are denoted.

<https://doi.org/10.1371/journal.pgen.1008073.g002>

17 instances of overlapping microbial and BA QTL on 12 chromosomes (LOD  $\geq 6.1$ ;  $P \leq 0.5$ ). This QTL overlap indicates there might be QTL with pleiotropic effects on BAs and the

microbiota, suggest that genetic variation influencing host BA profiles has an effect on compositional features of the gut microbiota, or genetic-driven variation in microbiota composition alters BAs. Examples of notable instances of overlapping bacterial and BA QTL, including *Akkermansia muciniphila* and Peptostreptococcaceae family are discussed in the Supporting Information (S1 Data).

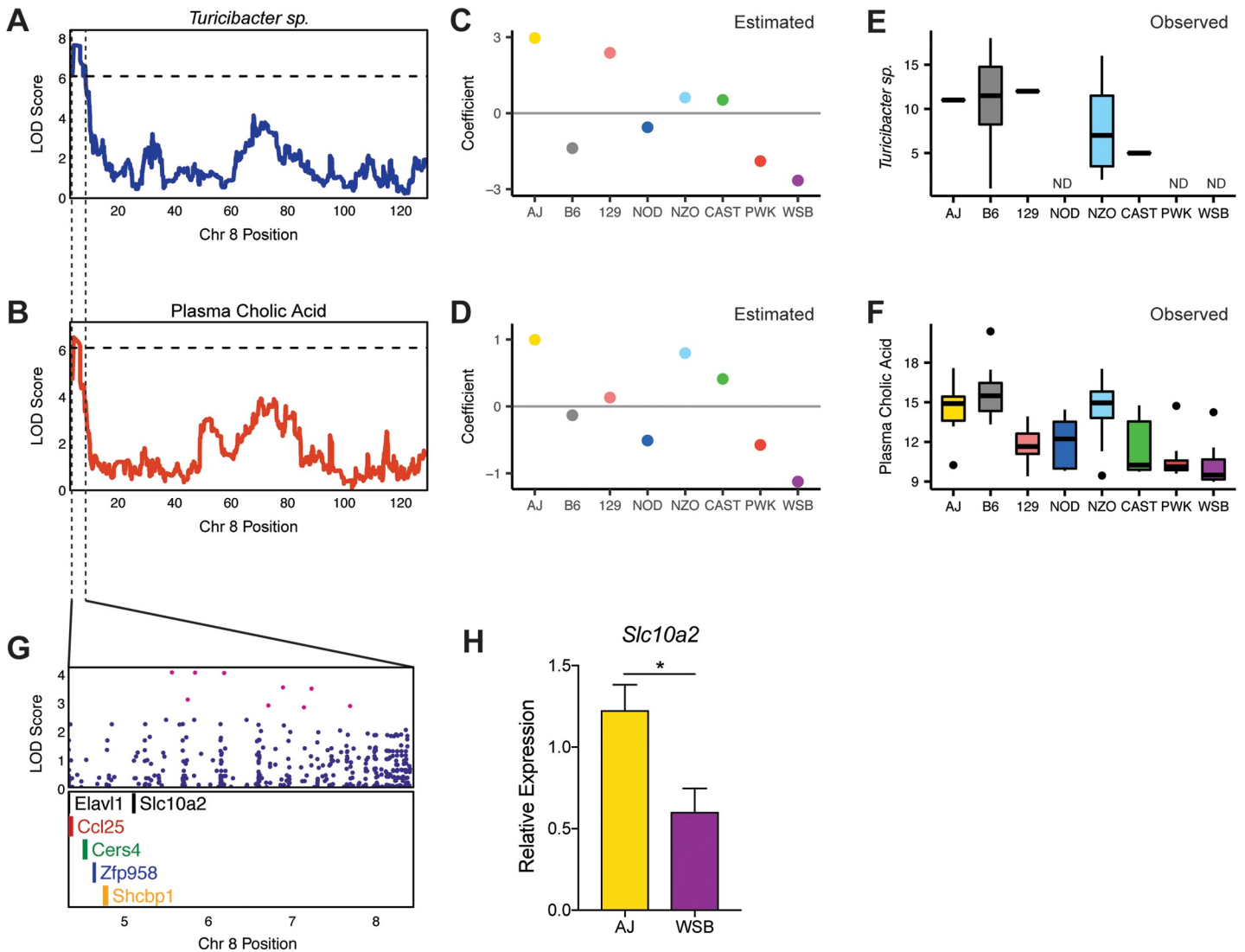
We focused our co-mapping analysis on chr 8 at ~ 5.5 Mbp, where *Turicibacter sp.* QTL and plasma cholic acid (CA) QTL overlap (Fig 3A and 3B). These traits were particularly interesting because both have been shown to be influenced by host genetics by previous studies. *Turicibacter* has been identified as highly heritable in both mouse and human genetic studies [24,27,45,50], and multiple reports have found differences in CA levels as a function of host genotype [18,46]. Furthermore, CA levels are influenced by both host genetics and microbial metabolism since it is synthesized by host liver enzymes from cholesterol and subsequently modified by gut microbes in the intestine. Notably, these co-mapping traits also share the same allele effects pattern, where the A/J and WSB haplotypes have strong positive and negative associations, respectively (Fig 3C and 3D).

To assess whether the trait patterns observed in the DO founder strains correspond to the observed allelic effects in the QTL mapping, we performed a separate characterization of the fecal microbiota composition and plasma bile acids in age-matched A/J and WSB animals fed the HF/HS diet. The founder strain allele patterns inferred from the QTL mapping closely resembled the observed levels of *Turicibacter sp.* (Fig 3E) and plasma CA in the founder strains (Fig 3F), where A/J animals had significantly higher levels of *Turicibacter sp.* and CA than WSB animals. However, *Turicibacter* levels in the founder strains do not completely mirror the estimated allele effects. This may be due to other genetic factors that also influence *Turicibacter* levels, as this taxa may be influenced by multiple host genes and levels of *Turicibacter* have previously been associated on chr 7 [24], 9 and 11 [50] in mice. Furthermore, *Turicibacter* and plasma CA were positively correlated in the DO mice ( $r = 0.43$ ,  $p = 3.53e^{-10}$ ). This finding is consistent with a previous study that found positive correlations between *Turicibacter* and unconjugated cecal BAs [62]. Taken together, the overlap between the *Turicibacter sp.* QTL and plasma CA QTL, along with the similar allele effects pattern, which reflect the values observed in the founder strains, provide strong evidence that these traits are related and they are responding to the common genetic driver.

### ***Slc10a2* is a candidate gene for *Turicibacter sp.* and plasma cholic acid**

We searched in the QTL confidence interval for candidate genes via high-resolution association mapping on chr 8 and identified SNPs associated with both microbial and BA traits. Among these we identified SNPs upstream of the gene *Slc10a2*, which encodes for the apical sodium-bile transporter (Fig 3G). *Slc10a2* is responsible for ~95% of BA reabsorption in the distal ileum and plays a key role in BA homeostasis [63]. In humans, mutations in this gene are responsible for primary BA malabsorption, resulting in interruption of enterohepatic circulation of BAs and decreased plasma cholesterol levels [64]. Likewise, *Slc10a2*<sup>-/-</sup> mice have a reduced total BA pool size, increased fecal BA concentrations and reduced total plasma cholesterol in comparison to wild-type mice [63]. Additionally, a comparison between germ-free and conventionally-raised mice found that expression of *Slc10a2* is downregulated in presence of the gut microbiota, suggesting microbes may influence the expression of the transporter [41].

Our analysis identified SNPs associated with levels of *Turicibacter sp.* and plasma CA at the QTL peak (Fig 3G). The SNPs with the strongest associations were attributed to the WSB and A/J haplotypes and fell on intergenic regions near *Slc10a2*. There is growing evidence that



**Fig 3. Co-mapping of *Turicibacter sp.* and plasma cholic acid (CA) QTL on chromosome 8.** Association of (A) fecal abundance of *Turicibacter sp.* and (B) plasma CA levels on chromosome (chr) 8. The x-axis indicates the position in Mbp along chr 8. The y-axis for the top panel and the y-axis in the bottom panel is the LOD score. Dashed line corresponds to LOD = 6.11 ( $P < 0.5$ ). A/J and WSB founder alleles are associated with higher and lower levels of *Turicibacter* and plasma CA levels, respectively. Estimated founder strain levels of *Turicibacter sp.* and plasma cholic acid were inferred in the DO population from the founder strain coefficients observed at the corresponding QTL on chr 8. The estimated founder strain abundance of (C) *Turicibacter* and (D) levels of plasma CA in the DO population reflects measured values observed in founder strains for (E) the abundance of *Turicibacter sp.* and (F) plasma cholic acid levels ( $n = 8$  mice/genotype, 4 male and 4 female). (G) SNPs (top panel) and protein coding genes (bottom panel) under the QTL interval. Magenta dots correspond to SNPs with the strongest association where the LOD drop  $< 1.5$  from the top SNP. (H) Relative expression of *Slc10a2* measured in the distal ileum by qRT-PCR in A/J and WSB parental strains ( $n = 6$ , 3 male and 3 female). Data are presented as mean  $\pm$  SEM; Welch's  $t$  test; \*  $p < 0.05$ . Correlation p-values adjusted for multiple tests using Benjamini and Hochberg correction. ND—not detected.

<https://doi.org/10.1371/journal.pgen.1008073.g003>

non-coding intergenic SNPs are often located in or closely linked to regulatory regions, suggesting that they may influence host regulatory elements and alter gene expression [65,66]. To assess if candidate gene expression patterns in the DO founders corresponds to the estimated allelic effects in the QTL mapping, we quantified *Slc10a2* expression in distal ileum samples from A/J and WSB mice by quantitative reverse transcriptase PCR (qRT-PCR). A/J mice exhibited significantly higher expression of *Slc10a2* compared to WSB mice (Fig 3H), which is consistent with estimated allele patterns for the overlapping *Turicibacter* and plasma CA QTLs



on chr 8 (Fig 3A and 3B). Remarkably, several studies have noted concomitant changes in microbiota composition and *Slc10a2* mRNA levels [67–69].

### A common genetic driver controls *Turicibacter sp.* and plasma cholic acid

We mapped QTL for *Turicibacter sp.* and for plasma CA levels to a common locus on chr 8 at 5–7 Mbp. Since the LOD profiles and allelic effects are highly similar, the QTL may be due to a single shared locus (pleiotropy) or multiple closely linked loci. We examined this question using a likelihood ratio testing of the null hypothesis of pleiotropy versus the alternative of two independent genetic regulators of these traits [70]. Analysis of 1000 bootstrap samples resulted in a p-value of 0.531, which is consistent with the presence of a single pleiotropic locus that affects both traits.

We next sought to understand the causal relationships between the microbe and the BA. We asked whether the relationship between the microbe and BA was causal, reactive or independent. To establish the directionality of the relationship, we applied mediation analysis where we conditioned one trait on the other [71]. When we conditioned *Turicibacter sp.* on plasma CA (QTL → BA → Microbe), we observed a LOD drop of 3.2 (Fig 4A and 4B). Likewise, when we conditioned the plasma cholic acid on the microbe (QTL → Microbe → BA) there was a LOD drop of 3.32 (Fig 4C and 4D). The partial mediation seen in both models suggests that the relationship between the microbe and the BA could be bidirectional, where they exert an effect on one another.

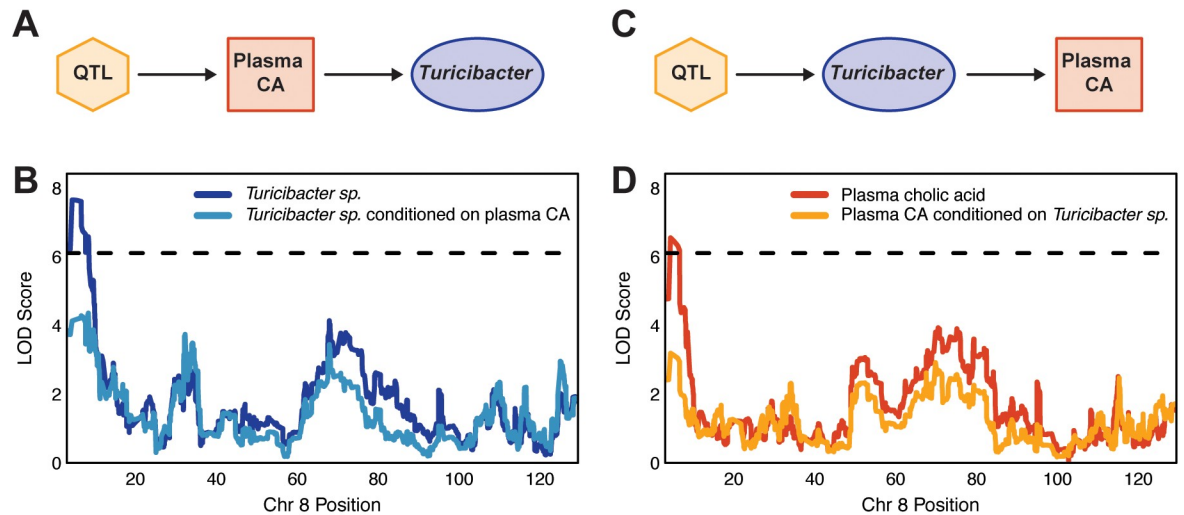
From this analysis, we can hypothesize this relationship can be explained by a pleiotropic model, where a single locus influences a microbial and a BA trait, and the microbial trait is also reactive to changes in the BA trait. It is important to note that statistical inference only partially explains the relationship between the traits and there may be other hidden variables that may further explain the relationship. The complex relationship depicted by the causal inference testing is consistent with the interplay between gut microbes and BAs in the intestine and their known ability to influence the other.

### Bile acids inhibit *Turicibacter sanguinis* growth at physiologically relevant concentrations

Due to the strong correlative relationship between the QTL, we tested whether there was a direct interaction between bile acids and *Turicibacter*. *Turicibacter* inhabits the small intestine where BAs are secreted upon consumption of a meal [72,73]. We screened the human isolate *Turicibacter sanguinis* for deconjugation and transformation activity *in vitro* by HPLC/MS-MS. We found that *T. sanguinis* deconjugated ~96–100% of taurocholic acid and glycochenodeoxycholic acid (Fig 5A) within 24 hours. It also transformed ~6 and 8% of CA and CDCA to 7-dHCA and 7-ketolithocholic acid (7-KLCA), respectively (Fig 5B and 5C). Both of these transformations require the action of the bacterial 7 $\alpha$ -hydroxysteroid dehydrogenase.

Based on these results, we asked if conjugated and unconjugated bile acids differentially modulate *T. sanguinis* growth. BA concentrations range from ~1–10 mM along the small intestine [74] to ~0.2–1 mM in the cecum [75]. Therefore, we grew *T. sanguinis* in the presence of either conjugated or unconjugated bile acids at physiologically relevant concentrations ranging from 0.1–5 mM. *T. sanguinis* growth decreased with increasing concentrations of BAs and growth was completely inhibited at 1 mM for unconjugated BAs and 5 mM for conjugated BAs (Fig 5D and 5E). Growth rate was significantly slower in the presence of 1 mM conjugated and 0.5mM unconjugated bile acids (Fig 5F). These results suggest that levels of BAs may affect abundance of *Turicibacter* in the gut.

To compare *T. sanguinis* sensitivity to conjugated bile acids relative to other small intestine colonizers, we grew four taxa (*Bacteroides thetaiotaomicron*, *Clostridium asparagiforme*,



**Fig 4. Mediation analysis and causal inference testing suggest causal relationship between *Turicibacter sp.* abundance and plasma cholic acid (CA) levels.** (A) Hypothetical causal model that proposes that cholic acid (CA) mediates the changes in *Turicibacter sp.* abundance. (B) Change in LOD score of plasma CA when adjusting for *Turicibacter sp.* abundance. The x-axis indicates the position in Mbp along chr 8. (C) Hypothetical causal model that proposes that *Turicibacter sp.* mediates changes in abundance of plasma CA levels. (D) Change in LOD score of *Turicibacter sp.* when controlling for plasma CA levels. Dashed lines correspond to LOD = 6.11 ( $p < 0.5$ ).

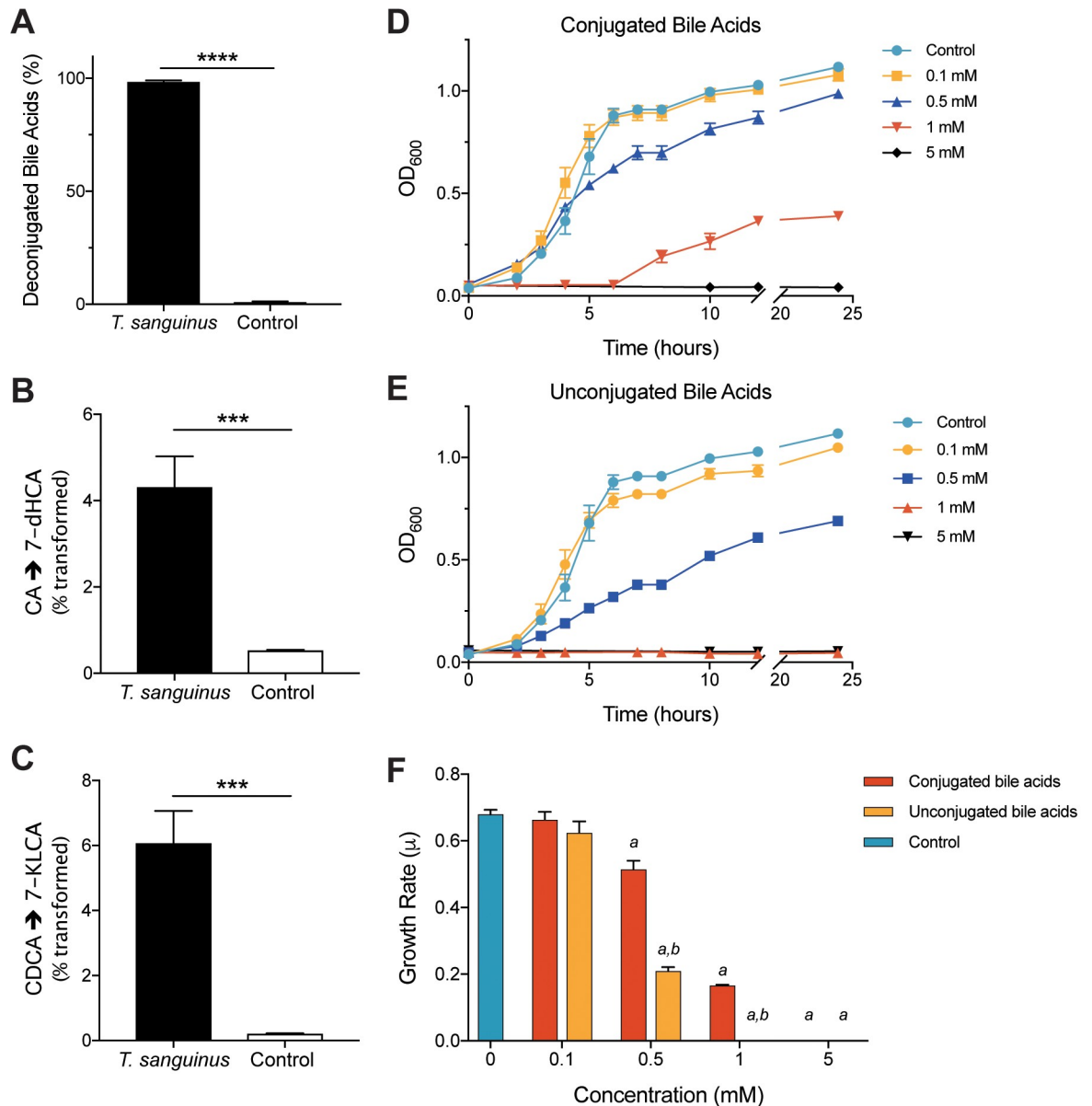
<https://doi.org/10.1371/journal.pgen.1008073.g004>

*Lactobacillus reuteri* and *Escherichia coli* MS200-1) known to colonize this region of the intestine with or without 1 mM conjugated bile acids. Members of these genera are known to have bile salt hydrolase (BSH) activity to deconjugate bile acids [35]. Unlike *T. sanguinis*, the addition of high levels of conjugated bile acids had little to no effect on the growth of these four gut microbes (S4 Fig). Consistent with these findings, *Turicibacter* abundance was negatively correlated with cecal TCA levels in the DO mice ( $r = -0.262$ ,  $p = 0.0035$ ).

Taken together, these data indicate that *T. sanguinis* is sensitive to higher concentrations of BA compared to other small intestine colonizers. These reciprocal effects between the BA and the bacterium provide biological evidence for the correlative relationship shown by the causal model testing. In summary, using a genetic approach, we identified and provide validation of a relationship between a genetic locus containing the BA transporter *Slc10a2*, and levels of *Turicibacter* and plasma cholic acid. Based on our findings, we hypothesize that the identified locus regulates expression of *Slc10a2*, altering active BA reabsorption in the ileum, leading to increased intestinal BA concentrations and alterations in the intestinal BA environment. Consequently, the resulting environmental change provides an unfavorable habitat for *Turicibacter*. In turn, lower levels of *Turicibacter* BA deconjugation activity leads to a decrease in circulating free plasma cholic acid levels.

## Conclusion

In this study, we performed the first known genetic mapping integration of gut microbiome and BA profiles. Using DO mice, we identified multiple QTL for gut microbes and bile acids spanning the host genome. These included loci that associated with individual microbial and BA traits, as well as loci with potential pleiotropic effects, where a single genetic region influenced both the abundance of a gut microbe and levels of a BA. While several studies suggest that host genetic variation has a minor impact on microbiota composition, there are overlapping findings among different studies in both human and mouse populations that indicate that specific bacterial taxa are influenced by host genetics. Our results in the DO population



**Fig 5. *Turicibacter sanguinis* and bile acid interactions.** (A) Percent of conjugated bile acids detected after 24-hour incubation with or without the presence of *T. sanguinis*. (B) Transformation of cholic acid (CA) to 7-dehydrocholic acid (7-dHCA), and (C) chenodeoxy cholic acid (CDCA) to 7-ketolithocholic acid (7-KLCA) by *T. sanguinis* after 24 hours. Growth of *T. sanguinis* in the presence of 0.1 mM, 0.5 mM, 1 mM and 5 mM (D) conjugated (equimolar pool of taurocholic acid (TCA) and glycochenodeoxycholic acid (GCDCA)), and (E) unconjugated (equimolar pool of cholic CA, CDCA, and deoxycholic acid (DCA)) bile acids over 24 hours. (F) Growth rate ( $\mu$ ) of *T. sanguinis* in medium supplemented with varying concentrations of conjugated and unconjugated bile acids. Data shown are from one experiment with three technical replicates. Data are presented as mean  $\pm$  SEM; one-way ANOVA followed by Tukey's multiple comparisons test; \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ .

<https://doi.org/10.1371/journal.pgen.1008073.g005>

corroborate several of these key findings (discussed in S1 Data). *Turicibacter sp.* is among the microbes consistently associated with host genetics. This work plus data from previous reports suggest that alterations in the BA pool driven by *Slc10a2* genetic variation and concomitant changes in expression/activity elicit an impact on gut microbiota community structure and influence the ability of *Turicibacter* to colonize and persist in the intestine. Although this

microbe deconjugates primary BAs, we found that it is also sensitive to elevated concentrations of both conjugated and unconjugated BAs. Future experiments are needed to examine how a decrease in *Slc10a2* expression changes intestinal BA profiles and the consequences on *Turicibacter* colonization. Additionally, this work identified multiple host-microbe-metabolite interactions that need to be validated with additional molecular studies. More broadly, our work demonstrates the power of genetics to identify novel interactions between microbial and metabolite traits and provides new testable hypotheses to further dissect factors that shape gut microbiota composition.

## Materials and methods

### Ethics statement

Animal care and study protocols were approved by the University of Wisconsin-Madison Animal Care and Use Committee (A005821) and were in compliance with all NIH animal welfare guidelines.

### Animals and sample collection

Animal care and study protocols were approved by the University of Wisconsin-Madison Animal Care and Use Committee. DO mice were obtained from the Jackson Laboratories (Bar Harbor, ME, USA) at ~4 weeks of age and maintained in the Department of Biochemistry vivarium at the University of Wisconsin-Madison. Mice were housed on a 12-hour light:dark cycle under temperature- and humidity-controlled conditions. Five waves of 100 DO mice each from generations, 17, 18, 19, 21, and 23 were obtained at intervals of 3–6 months. Each wave was composed of equal numbers of male and female mice. All mice were fed a high-fat high-sucrose diet (TD.08811, Envigo Teklad, 44.6% kcal fat, 34% carbohydrate, and 17.3% protein) *ad libitum* upon arrival to the facility. Mice were kept in the same vivarium room and were individually housed to monitor food intake and prevent coprophagy between animals. DO mice were sacrificed at 22–25 weeks of age.

The eight DO founder strains (C57BL/6J, A/J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HILtJ, PWK/PhJ, WSB/EiJ and CAST/EiJ) were obtained from the Jackson Laboratories. Mice were bred at the University of Wisconsin-Madison Biochemistry Department. Mice were housed by strain and sex (2–5 mice/cage), with the exception of CAST that required individual housing. Inbred founder mice were housed under the same environmental conditions as the DO animals. Like the DO mice, the eight founder strains were maintained on the HF/HS diet and were sacrificed at 22 weeks of age, except for NZO males that were sacrificed at 14 weeks, due to high mortality attributable to severe disease.

For both DO and founder mice, fecal samples for 16S rRNA sequencing were collected immediately before sacrifice after a 4 hour fast. Cecal contents, plasma, and additional tissues were harvested promptly after sacrifice and all samples were immediately flash frozen in liquid nitrogen and stored at -80°C until further processing.

### DNA extraction

DNA was isolated from feces using a bead-beating protocol [18]. Mouse feces (~1 pellet per animal) were re-suspended in a solution containing 500µl of extraction buffer [200mM Tris (pH 8.0), 200mM NaCl, 20mM EDTA], 210µl of 20% SDS, 500µl phenol:chloroform:isoamyl alcohol (pH 7.9, 25:24:1) and 500µl of 0.1-mm diameter zirconia/silica beads. Cells were mechanically disrupted using a bead beater (BioSpec Products, Barlesville, OK; maximum setting for 3 min at room temperature), followed by extraction with phenol:chloroform:isoamyl

alcohol and precipitation with isopropanol. Contaminants were removed using QIAquick 96-well PCR Purification Kit (Qiagen, Germantown, MD, USA). Isolated DNA was eluted in 5 mM Tris/HCL (pH 8.5) and was stored at  $-80^{\circ}\text{C}$  until further use.

### 16S rRNA sequencing

PCR was performed using universal primers flanking the variable 4 (V4) region of the bacterial 16S rRNA gene [76]. Genomic DNA samples were amplified in duplicate. Each reaction contained 10–30 ng genomic DNA, 10  $\mu\text{M}$  each primer, 12.5  $\mu\text{l}$  2x HiFi HotStart ReadyMix (KAPA Biosystems, Wilmington, MA, USA), and water to a final reaction volume of 25  $\mu\text{l}$ . PCR was carried out under the following conditions: initial denaturation for 3 min at  $95^{\circ}\text{C}$ , followed by 25 cycles of denaturation for 30 s at  $95^{\circ}\text{C}$ , annealing for 30 s at  $55^{\circ}\text{C}$  and elongation for 30 s at  $72^{\circ}\text{C}$ , and a final elongation step for 5 min at  $72^{\circ}\text{C}$ . PCR products were purified with the QIAquick 96-well PCR Purification Kit (Qiagen, Germantown, MD, USA) and quantified using Qubit dsDNA HS Assay kit (Invitrogen, Oregon, USA). Samples were equimolar pooled and sequenced by the University of Wisconsin–Madison Biotechnology Center with the MiSeq 2x250 v2 kit (Illumina, San Diego, CA, USA) using custom sequencing primers.

### 16S analysis

Demultiplexed paired end fastq files generated by CASAVA (Illumina) and a mapping file were used as input files. Sequences were processed, quality filtered and analyzed with QIIME2 (version 2018.4) (<https://qiime2.org>), a plugin-based microbiome analysis platform [77]. DADA2 [48] was used to denoise sequencing reads with the q2-dada2 plugin for quality filtering and identification of *de novo* exact sequence variants (ESVs) (i.e. 100% exact sequence match). This resulted in 20,831,573 total sequences with an average of 52,078 sequences per sample for the DO mice, and 2,128,796 total sequences with an average of 34,335.4 sequences per sample for the eight DO founder strains. Sequence variants were aligned with mafft [78] with the q2-alignment plugin. The q2-phylogeny plugin was used for phylogenetic reconstruction via FastTree [79]. Taxonomic classification was assigned using classify-sklearn [80] against the Greengenes 13\_8 99% reference sequences [81]. Alpha- and beta-diversity (weighted and unweighted UniFrac [82] analyses were performed using q2-diversity plugin at a rarefaction depth of 10000 sequences per sample. For the DO mice, one sample (DO071) was removed from subsequent analysis because it did not reach this sequencing depth. For analysis of the eight DO founder strains, one sample (NOD5) was removed because it did not reach this sequencing depth. Subsequent processing and analysis were performed in R (v.3.5.1), and data generated in QIIME2 was imported into R using Phyloseq [83]. Sequencing data was normalized by cumulative sum scaling (CSS) using MetagenomeSeq [84]. Summaries of the taxonomic distributions were generated by collapsing normalized ESV counts into higher taxonomic levels (genus to phylum) by phylogeny. We defined a core measurable microbiota (CMM) [24] to include only microbial traits present in 20% of individuals in the QTL mapping. In total, 86 ESVs and 42 collapsed microbial taxonomies comprised the CMM.

### Sample preparation for plasma bile acid analysis

40  $\mu\text{L}$  of DO plasma collected at sacrifice (30  $\mu\text{L}$  used for founder strains) were aliquoted into a tube with 10  $\mu\text{L}$  SPLASH Lipidomix internal standard mixture (Avanti Polar Lipids, Inc.). Protein was precipitated by addition of 215  $\mu\text{L}$  MeOH. After the mixture was vortexed for 10 s, 750  $\mu\text{L}$  methyl tert-butyl ether (MTBE) were added as extraction solvent and the mixture was vortexed for 10 s and mixed on an orbital shaker for 6 min. Phase separation was induced by adding 187.5  $\mu\text{L}$  of water followed by 20 s of vortexing. All steps were performed at  $4^{\circ}\text{C}$  on ice.

Finally, the mixture was centrifuged for 4 min at 14,000 x g at 4°C and stored at -80°C. For targeted bile acids analysis, samples were thawed on ice. 400 µL of ethanol were added to further precipitate protein, as well as 15 µL of isotope-labeled internal standard mix (12.5 µM d4-TαMCA, 10 µM d4-CDCA). The samples were vortexed for 20 s and centrifuged for 4 min at 14,000 g at 4°C after which the supernatant (ca. 1000 µL) was taken out and dried down. Dried supernatants were resuspended in 60 µL mobile phase (50%B), vortexed for 20 s, centrifuged for 4 min at 14,000 g and then 50 µL were transferred to vials with glass inserts for MS analysis.

### Sample preparation for cecal bile acid analysis

30 ± 7.5 mg cecal contents along with 10 µL SPLASH Lipidomix internal standard mixture were aliquoted into a tube with a metal bead and 270 µL MeOH were added for protein precipitation. To each tube, 900 µL MTBE and 225 µL of water were added as extraction solvents. All steps were performed at 4°C on ice. The mixture was homogenized by bead beating for 8 min at 25 Hz. Finally, the mixture was centrifuged for 4–8 min at 11,000 x g at 4°C. Subsequent processing for the DO mice and eight DO founder strains differed due to other analyses performed on the samples that are not presented in this paper. For DO samples, 100 µL of the aqueous and 720 µL of organic layer were combined and stored at -80°C. For analysis, these were thawed on ice and 400 µL of ethanol were added to further precipitate protein, as well as 15 µL of isotope-labeled internal standard mix (12.5 µM d4-TαMCA, 10 µM d4-CDCA). The samples were vortexed for 20 s and centrifuged for 4 min at 14,000 g at 4°C after which the supernatant (ca. 1000 µL) was taken out and dried down. Dried supernatants were resuspended in 100 µL mobile phase (50%B), vortexed for 20 s, centrifuged for 8 min at 14,000 g and then 50 µL were transferred to vials with glass inserts for MS analysis. For the eight DO founder strains, the mixture was dried down including all solid parts and stored dried at -80°C. For targeted bile acid analysis, these dried down samples were then thawed on ice and reconstituted in 270 µL of methanol, 900 µL of MTBE, and 225 µL of water. 400 µL of ethanol were added to further precipitate protein, as well as 15 µL of isotope-labeled internal standard mix (12.5 µM d4-TαMCA, 10 µM d4-CDCA). The mixture was bead beat for 8 min at 25 Hz and centrifuged at 14,000 g for 8 minutes after which the supernatant (ca. 1500 µL) was taken out and dried down. Dried supernatants were resuspended in 100 µL mobile phase (50%B), vortexed for 20 s, centrifuged for 4 min at 14,000 g and then 90 µL were transferred to vials with glass inserts for MS analysis.

### Measurement and analysis of mouse bile acids

LC-MS analysis was performed in randomized order using an Acquity CSH C18 column held at 50°C (100 mm × 2.1 mm × 1.7 µm particle size; Waters) connected to an Ultimate 3000 Binary Pump (400 µL/min flow rate; Thermo Scientific). Mobile phase A consisted of 10 mM ammonium acetate containing 1 mL/L ammonium hydroxide. Mobile phase B consisted of MeOH with the same additives [85]. Mobile phase B was initially held at 50% for 1.5 min and then increased to 70% over 13.5 min. Mobile phase B was further increased to 99% over 0.5 min and held for 2.5 min. The column was re-equilibrated for 5.5 min before the next injection. Twenty microliters of plasma sample or ten microliters of cecum sample were injected by an Ultimate 3000 autosampler (Thermo Scientific). The LC system was coupled to a TSQ Quantiva Triple Quadrupole mass spectrometer (Thermo Scientific) by a heated ESI source kept at 325°C (Thermo Scientific). The inlet capillary was kept at 350°C, sheath gas was set to 15 units, auxiliary gas to 10 units, and the negative spray voltage was set to 2,500 V. For targeted analysis the MS was operated in negative single reaction monitoring (SRM) mode

acquiring scheduled, targeted scans to quantify selected bile acid transitions, with two transitions for each species' precursor and 3 min retention time windows. Collision energies were optimized for each species and ranging from 20–55 V. Due to insufficient fragmentation for unconjugated bile acids, the precursor was monitored as one transition with a CE of 20 V. MS acquisition parameters were 0.7 FWHM resolution for Q1 and Q3, 1 s cycle time, 1.5 mTorr CID gas and 3 s Chrom filter. In total, 27 bile acids, including 14 unconjugated, 9 tauro- and 4 glycine-conjugated species, were measured. The resulting bile acid data were processed using Skyline 3.6.0.10493 (University of Washington). For each species, one transition was picked for quantitation, while the other was used for retention time confirmation. Normalization of the quantitative data was performed to the internal standard d4-CDCA as indicated in Eq 1.

$$(\text{Peak Area}/\text{d4-CDCA Peak Area}) \cdot \text{Average of d4-CDCA Peak Area} \quad \text{Eq 1}$$

## Genotyping

Genotyping was performed on tail biopsies as previously described [42] using the Mouse Universal Genotyping Array (GigaMUGA; 143,259 markers) [86] at Neogen (Lincoln, NE). Genotypes were converted to founder strain-haplotype reconstructions using a hidden Markov model (HMM) implemented in the R/qlt2 package [49]. We interpolated the GigaMUGA markers onto an evenly spaced grid with 0.02-cM spacing and added markers to fill in regions with sparse physical representation, which resulted in 69,005 pseudomarkers.

## QTL mapping

We performed QTL mapping using the R package R/qlt2 [49]. QTL mapping was done through a regression of the phenotype on the founder haplotype probabilities estimated with an HMM designed for multi-parental populations. Genome scans were performed for each phenotype with sex, cohort (wave), and days on diet included as additive covariates. Genetic similarity between mice was accounted for using a kinship matrix based on the leave-one-chromosome-out (LOCO) methods [87]. For microbial QTL mapping, normalized gut microbiota abundance data transformed to normal quantiles. For bile acid QTL mapping, normalized plasma and cecal bile acid levels were log<sub>2</sub> transformed. The mapping statistic reported is the log<sub>10</sub> likelihood ratio (LOD score). The QTL support interval was defined using the 95% Bayesian confidence interval. Significant and suggestive QTL were determined at a genome-wide threshold of  $P \leq 0.05$  ( $\text{LOD} \geq 7.66$ ) and  $P \leq 0.2$  ( $\text{LOD} \geq 6.80$ ), respectively. We used a common significance threshold for all phenotypes, by pooling the permutation results for the individual phenotypes. No adjustment was made for the search across multiple phenotypes.

## Mediation/Pleiotropy analysis

To assess whether two co-mapping traits were caused by a pleiotropic locus, we used a likelihood ratio test implemented with the open source R package R/qlt2pleio [70]. Here, we compared the alternative hypothesis of two distinct loci with the null hypothesis of pleiotropy for two traits that map to the same genetic region. Parametric bootstrapping was used to determine statistical significance. Mediation analysis was applied to identify whether a microbe or bile acid were likely to be a causal mediator of the QTL as presented in Li et al. [88]. This analysis was adapted from a general approach previously described to differentiate target from mediator variables [89]. The effect of a mediator on a target was evaluated by performing an allele scan or SNP scan using the target adjusted by mediator. Only individuals with both values for both traits were considered for mediation analysis. Traits with a LOD drop >2 after controlling for the mediator were considered for further causality testing. To statistically assess

causality between microbial and bile acid trait sets (causal, reactive, independent, undecided), a causal model selection test [90] was applied using the R packages R/intermediate and R/qlt2. Causal model selection tests were evaluated on both alleles and SNPs in peak region.

### RNA extraction

Total RNA was extracted from flash-frozen distal ileum tissues by TRIzol extraction and further cleaned using the RNeasy Mini Kit (Qiagen, Germantown, MD, USA). DNA was removed by on-column DNase digestion (Qiagen). Purified RNA was quantified using a Nanodrop 2000 spectrophotometer.

### Quantitative Real-Time PCR

SuperScript II Reverse Transcriptase with oligo(dT) primer (all from Invitrogen, Carlsbad, CA, USA) was used to synthesize 20  $\mu$ l cDNA templates from 1  $\mu$ g purified RNA. cDNA was diluted 2X before use and qRT-PCR reactions were prepared in a 10  $\mu$ l volume using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, USA) and 400 nM specific primers targeting the gene of interest (SLC10A2-F [5'- TGGGTTTCTTCCTGGCTAGACT-3']; SLC10A2-R [5'- TGTTCCTGCATTCCAGTTCCAA-3'] [91]). All reactions were performed in triplicate. Reactions were run on a CFX96 Real-Time PCR System (Bio-Rad, Hercules, CA, USA). The  $2^{-\Delta\Delta C_t}$  method [92] was used to calculate relative changes in gene expression and all results were normalized to GAPDH.

### Bacterial culturing

Bacterial strains were obtained from DSMZ and ATCC. All strains were cultured at 37°C under anaerobic conditions using an anaerobic chamber (Coy Laboratory Products) with a gas mix of 5% hydrogen, 20% carbon dioxide and 75% nitrogen. Strains were grown in rich medium (S5 Table) that was filter sterilized and stored in the anaerobic chamber at least 24 hours prior to use. *L. reuteri* was grown in medium supplemented with 20 mM glucose. For all *in vitro* assays, cultures used for inoculation were grown overnight at 37°C in 10 mL 14b medium in anaerobic Hungate tubes. Stock solutions of conjugated bile acids (TCA, GCDCA) and unconjugated bile acids (CA, CDCA, DCA) were prepared to a final concentration of 100 mM and used for all *in vitro* assays. All bile acids used were soluble in methanol.

### Microbial bile acid metabolism screen

Stock solutions of conjugated and unconjugated bile acids (100 mM) were added to 3 ml 14b medium to obtain a final concentration of 100  $\mu$ M total bile acid. Tubes were inoculated with a *T. sanguinis* cultured overnight, then incubated in the anaerobic chamber at 37°C for 48 hours. At the 24- and 48-hour timepoints, 1 mL of each culture was removed and the supernatant was collected after brief centrifugation. Each culture supernatant was diluted 10x in initial running solvent (30:70 MeOH:10 mM ammonium acetate). Samples were spun at max speed for 3 minutes to remove suspended particles prior to loading on the uHPLC. Samples were analyzed using a uHPLC coupled with a high-resolution mass spectrometer.

### Microbial bile acid screen uHPLC-MS/MS parameters

10  $\mu$ L aliquots of diluted supernatant samples were analyzed using a uHPLC-MS/MS system consisting of a Vanquish uHPLC coupled by electrospray ionization (ESI) (negative mode) to a hybrid quadrupole-high-resolution mass spectrometer (Q Exactive Orbitrap; Thermo Scientific). Liquid chromatography separation was achieved on an Acquity UPLC BEH C<sub>18</sub> column



(2.1-by 100-mm column, 1.7- $\mu$ m particle size) heated to 50°C. Solvent A was 10 mM Ammonium acetate, pH 6; solvent B was 100% methanol. The total run time was 31.5 minutes with the following gradient: 0 min, 30% B; 0.5 min, 30% B; 24 min, 100% B; 29 min, 100% B; 29 min, 30% B; 31.5 min, 30% B. Bile acid peaks were identified using the Metabolomics Analysis and Visualization Engine (MAVEN) [93].

### Growth curves

Bacterial growth rate was measured in medium 14b supplemented with either 100  $\mu$ M, 300  $\mu$ M, 1 mM bile acids or methanol control. Medium was dispensed inside an anaerobic chamber into Hungate tubes. Tubes containing 10 mL of medium were inoculated with 30  $\mu$ L of an overnight culture and incubated at 37°C for 24 hours. *T. sanguinis* was grown with shaking to disrupt the formation of flocculent colonies. Growth was monitored as the increase in absorbance at 600 nm in a Spectronic 20D+ spectrophotometer (Thermo Scientific, Waltham, MA, USA). Growth rate was determined as  $\mu = \ln(X/X_0)/T$ , where  $X$  is the OD<sub>600</sub> value during the linear portion of growth and  $T$  is time in hours. Values given are the mean  $\mu$  values from two independent cultures done in triplicate.

### Statistical analysis

All statistical analyses were performed in R (v.3.5.1) [94]. Unless otherwise indicated in the figure legends, differences between groups were evaluated using unpaired two-tailed Welch's t-test. For multiple comparisons, Kruskal-Wallis test was used if ANOVA conditions were not met, followed by Mann-Whitney/Wilcoxon rank-sum for multiple comparisons and adjusted for multiple testing using the Benjamini-Hochberg FDR procedure. The correlation between the abundance of microbial taxa was performed using Spearman's correlation in the "Hmisc" (v.4.1-1) R package [95]. The p-values were adjusted using the Benjamini and Hochberg method, and correlation coefficients were visualized using the "pheatmap" (v.1.0.10) [96]. Multiple groups were compared by Kruskal-Wallis test and adjusted for multiple testing using the Benjamini-Hochberg FDR procedure. Significance was determined as p-value < 0.05. To assess magnitude of variability of the CMMs, summary statistics were calculated on each CMM (taxa and ESVs). Non-parametric-based PERMANOVA statistical test [97] with 999 Monte Carlo permutations was used to compare microbiota compositions among groups using the Vegan R package [98].

### Supporting information

#### S1 Data. Extended results and discussion.

(PDF)

**S1 Fig. Principal coordinate analysis (PCoA) of unweighted UniFrac distances for fecal samples.** PCoA shows significant clustering by (A) sex ( $F = 5.572$ ,  $p = 0.001$ ) and (B) wave ( $F = 16.954$ ,  $p = 0.001$ ). Clustering by treatment evaluated by PERMANOVA.

(TIF)

**S2 Fig. Plasma and cecal bile acids group by sex, but not wave.** PCAs of plasma bile acid profiles colored by (A) sex ( $p < 0.0001$ ) and (B) wave ( $p = 0.594$ ), and PCAs of cecal bile acid profiles colored by (C) sex ( $p = 0.011$ ) and (D) wave ( $p = 0.207$ ). Kruskal Wallis one-way test followed by Wilcoxon pair-wise multiple comparisons with Benjamini and Hochberg correction.

(TIF)

**S3 Fig. Related bile acid species map associate to same locus.** (A) Haplotype effects and LOD scores of plasma taurodeoxycholic acid (TDCA) and (B) cecal deoxycholic acid (DCA). For each plot, the x-axis is the physical position in Mbp along chr 12. The y-axis for the top panel is the effect coefficient depicting the estimated contributions of each founder allele, and the y-axis in the bottom panel is the LOD score. (C) Cecal levels of isolithocholic acid (ILCA) and lithocholic acid (LCA) associate to same locus on chr 11. (D) Estimated founder allele effects for cecal ILCA and (E) LCA. (F) Genes under cecal LCA and ILCA QTL interval. Vertical dashed lines denote QTL confidence interval. Horizontal dashed lines correspond to  $\text{LOD} = 6.11$  ( $p < 0.5$ ).

(TIF)

**S4 Fig. Gut associated bacteria have differential growth responses to conjugated bile acids.**

Growth rate in the presence of 1 mM conjugated bile acids or methanol control for (A) *Bacteroides thetaiotaomicron*, (B) *Clostridium asparagiforme*, (C) *Escherichia coli* MS200-1, and (D) *Lactobacillus reuteri*. Data shown are from duplicate experiments with three technical replicates. Data are presented as mean  $\pm$  SEM; Welch's *t* test; no significant differences were observed between growth conditions for any of the tested organisms.

(TIF)

**S5 Fig. Peptostreptococcaceae and plasma bile acids co-map on chromosome (chr) 3.**

Haplotype effects and LOD scores of (A) Peptostreptococcaceae family, (B) plasma cholic acid (CA), (C) plasma chenodeoxycholic acid (CDCA), (D) plasma muricholic acid (MCA), (E) plasma ursodeoxycholic acid (UDCA), and (F) plasma 7-dehydrocholic acid (7-dHCA). For each plot, the x-axis is the physical position in Mbp along chr 3. The y-axis for the top panel is the effect coefficient depicting the estimated contributions of each founder allele, and the y-axis in the bottom panel is the LOD score. Horizontal dashed line corresponds to  $\text{LOD} = 5.5$ . All overlapping QTL have positive association with the NOD allele. (G) Protein coding genes under QTL interval.

(TIF)

**S6 Fig. Exact sequence variant of *Akkermansia muciniphila* and plasma bile acid QTL overlap on chromosome (chr) 1.**

Haplotype effects and LOD scores of (A) *A. muciniphila* (B) plasma cholic acid (CA), (C) plasma muricholic acid (MCA), and (D) plasma 7-dehydrocholic acid (7-dHCA). For each plot, the x-axis is the physical position in Mbp along chr 1. The y-axis for the top panel is the effect coefficient depicting the estimated contributions of each founder allele, and the y-axis in the bottom panel is the LOD score. Horizontal dashed line corresponds to  $\text{LOD} = 5.5$ . (E) Protein coding genes under 10 Mbp QTL interval. Spearman correlations in the DO mice between *A. muciniphila* and (F) plasma CA, (G) plasma MCA, and (H) plasma 7-dHCA levels. Correlation p-values adjusted for multiple tests using Benjamini and Hochberg correction. Higher levels of these microbial and bile acid traits were associated with the NZO haplotype and lower levels were associated with the 129 haplotype. (E) Protein coding genes under 10 Mbp QTL interval. Dashed lines denote QTL confidence interval.

Spearman correlations in the DO mice between *A. muciniphila* and (F) plasma CA, (G) plasma MCA, and (H) plasma 7-dHCA levels. Correlation p-values adjusted for multiple tests using Benjamini and Hochberg correction.

(TIF)

**S1 Table. Measures of variability of cecal and plasma bile acids in DO mice.** Bile acid levels are presented as  $\log_2(\text{peak area})$ ;  $n = 384$ ; SD, standard deviation.

(XLSX)

**S2 Table. Measures of variability of microbial exact sequence variants (ESVs) or taxon (phylum, class, order, family, genus) in DO mice.** Data presented as normalized read counts; n = 399; SD, standard deviation.

(XLSX)

**S3 Table. Correlations among microbial taxa, bile acid and weight traits.** Spearman's rank correlation. Only microbial exact sequence variants, genera and family included in figure. Correlations shown passed FDR < 0.01 cut-off and correlation coefficient either < -0.35 or > 0.35. Correlating bile acids from same tissue removed from table for brevity.

(XLSX)

**S4 Table. QTL peaks for gut microbiota, plasma and cecal bile acid, and weight traits in the Diversity Outbred mice.** Only QTL with LOD > 6.1 shown. "Pos" is peak position in Mbp. "ci\_lo" and "ci\_hi" correspond to the positions for the 95% bayesian confidence interval.

(XLSX)

**S5 Table. Media used for bacterial culture.** Medium 14(b) recipe.

(XLSX)

## Acknowledgments

The authors thank the University of Wisconsin Biotechnology Center DNA Sequencing Facility for providing sequencing and support services, and the University of Wisconsin Center for High Throughput Computing (CHTC) in the Department of Computer Sciences for providing computational resources, support, and assistance. We also thank Paul Dawson for his feedback.

## Author Contributions

**Conceptualization:** Julia H. Kemis, Mark P. Keller, Karl W. Broman, Alan D. Attie, Federico E. Rey.

**Data curation:** Julia H. Kemis, Lindsay L. Traeger, Brian S. Yandell, Karl W. Broman.

**Formal analysis:** Julia H. Kemis, Frederick J. Boehm, Brian S. Yandell, Karl W. Broman.

**Funding acquisition:** Mark P. Keller, Alan D. Attie, Federico E. Rey.

**Investigation:** Julia H. Kemis, Vanessa Linke, Kelsey L. Barrett, Frederick J. Boehm, Lindsay L. Traeger, Mary E. Rabaglia, Kathryn L. Schueler, Donald S. Stapleton.

**Methodology:** Brian S. Yandell, Karl W. Broman.

**Project administration:** Julia H. Kemis, Jason D. Russell.

**Resources:** Daniel M. Gatti, Gary A. Churchill, Daniel Amador-Noguez, Brian S. Yandell, Karl W. Broman, Joshua J. Coon.

**Software:** Julia H. Kemis, Daniel M. Gatti, Gary A. Churchill, Brian S. Yandell, Karl W. Broman.

**Supervision:** Mark P. Keller, Alan D. Attie, Federico E. Rey.

**Visualization:** Julia H. Kemis, Brian S. Yandell.

**Writing – original draft:** Julia H. Kemis, Federico E. Rey.

**Writing – review & editing:** Julia H. Kemis, Vanessa Linke, Kelsey L. Barrett, Frederick J. Boehm, Lindsay L. Traeger, Mark P. Keller, Gary A. Churchill, Brian S. Yandell, Karl W. Broman, Joshua J. Coon, Alan D. Attie, Federico E. Rey.

## References

1. Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on human health: an integrative view. *Cell*. 2012; 148: 1258–1270. <https://doi.org/10.1016/j.cell.2012.01.035> PMID: 22424233
2. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, et al. Richness of human gut microbiome correlates with metabolic markers. *Nature*. 2013; 500: 541–546. <https://doi.org/10.1038/nature12506> PMID: 23985870
3. Sommer F, Bäckhed F. The gut microbiota—masters of host development and physiology. *Nat Rev Microbiol*. 2013; 11: 227–238. <https://doi.org/10.1038/nrmicro2974> PMID: 23435359
4. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature*. 2012; 489: 220–230. <https://doi.org/10.1038/nature11550> PMID: 22972295
5. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science*. 2016; 352: 565–569. <https://doi.org/10.1126/science.aad3369> PMID: 27126040
6. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature*. 2018; 555: 210–215. <https://doi.org/10.1038/nature25973> PMID: 29489753
7. Dill-McFarland K, Tang Z-Z, Kemis J, Kerby R, Chen G, Palloni A, et al. Social relationships, social isolation, and the human gut microbiota. *bioRxiv*. 2018; 428938. <https://doi.org/10.1101/428938>
8. Ussar S, Fujisaka S, Kahn CR. Interactions between host genetics and gut microbiome in diabetes and metabolic syndrome. *Mol Metab*. 2016; 5: 795–803. <https://doi.org/10.1016/j.molmet.2016.07.004> PMID: 27617202
9. Hall AB, Tolonen AC, Xavier RJ. Human genetic variation and the gut microbiome in disease. *Nat Rev Genet*. 2017; 18: 690–699. <https://doi.org/10.1038/nrg.2017.63> PMID: 28824167
10. Karlsson F, Tremaroli V, Nielsen J, Backhed F. Assessing the Human Gut Microbiota in Metabolic Diseases. *Diabetes*. 2013; 62: 3341–3349. <https://doi.org/10.2337/db13-0844> PMID: 24065795
11. Petersen C, Round JL. Defining dysbiosis and its influence on host immunity and disease. *Cell Microbiol*. 2014; 16: 1024–1033. <https://doi.org/10.1111/cmi.12308> PMID: 24798552
12. Petra AI, Panagiotidou S, Hatzigelaki E, Stewart JM, Conti P, Theoharides TC. Gut-Microbiota-Brain Axis and Its Effect on Neuropsychiatric Disorders With Suspected Immune Dysregulation. *Clin Ther*. 2015; 37: 984–995. <https://doi.org/10.1016/j.clinthera.2015.04.002> PMID: 26046241
13. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekman R, et al. Human genetics shape the gut microbiome. *Cell*. 2014; 159: 789–799. <https://doi.org/10.1016/j.cell.2014.09.053> PMID: 25417156
14. Plovier H, Everard A, Druart C, Depommier C, Van Hul M, Geurts L, et al. A purified membrane protein from *Akkermansia muciniphila* or the pasteurized bacterium improves metabolism in obese and diabetic mice. *Nat Med*. 2017; 23: 107–113. <https://doi.org/10.1038/nm.4236> PMID: 27892954
15. Kasahara K, Krautkramer KA, Org E, Romano KA, Kerby RL, Vivas EI, et al. Interactions between *Roseburia intestinalis* and diet modulate atherogenesis in a murine model. *Nat Microbiol*. 2018; 1. <https://doi.org/10.1038/s41564-018-0272-x> PMID: 30397344
16. Kurilshikov A, Wijmenga C, Fu J, Zhernakova A. Host Genetics and Gut Microbiome: Challenges and Perspectives. *Trends Immunol*. 2017; 38: 633–647. <https://doi.org/10.1016/j.it.2017.06.003> PMID: 28669638
17. Parks BW, Nam E, Org E, Kostem E, Norheim F, Hui ST, et al. Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell Metab*. 2013; 17: 141–152. <https://doi.org/10.1016/j.cmet.2012.12.007> PMID: 23312289
18. Kreznar JH, Keller MP, Traeger LL, Rabaglia ME, Schueler KL, Stapleton DS, et al. Host Genotype and Gut Microbiome Modulate Insulin Secretion and Diet-Induced Metabolic Phenotypes. *Cell Rep*. 2017; 18: 1739–1750. <https://doi.org/10.1016/j.celrep.2017.01.062> PMID: 28199845
19. Vijay-Kumar M, Aitken JD, Carvalho FA, Cullender TC, Mwangi S, Srinivasan S, et al. Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science*. 2010; 328: 228–231. <https://doi.org/10.1126/science.1179721> PMID: 20203013

20. Henao-Mejia J, Elinav E, Jin C, Hao L, Mehal WZ, Strowig T, et al. Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. *Nature*. 2012; 482: 179–185. <https://doi.org/10.1038/nature10809> PMID: 22297845
21. Rehman A, Sina C, Gavrilova O, Hasler R, Ott S, Baines JF, et al. Nod2 is essential for temporal development of intestinal microbial communities. *Gut*. 2011; 60: 1354–1362. <https://doi.org/10.1136/gut.2010.216259> PMID: 21421666
22. Lamas B, Richard ML, Leducq V, Pham H-P, Michel M-L, Da Costa G, et al. CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands. *Nat Med*. 2016; 22: 598–605. <https://doi.org/10.1038/nm.4102> PMID: 27158904
23. Leamy LJ, Kelly SA, Nietfeldt J, Legge RM, Ma F, Hua K, et al. Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. *Genome Biol*. 2014; 15: 552. <https://doi.org/10.1186/s13059-014-0552-6> PMID: 25516416
24. Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J, et al. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci U S A*. 2010; 107: 18933–18938. <https://doi.org/10.1073/pnas.1007028107> PMID: 20937875
25. McKnite AM, Perez-Munoz ME, Lu L, Williams EG, Brewer S, Andreux PA, et al. Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. White BA, editor. *PLoS One*. 2012; 7: e39191. <https://doi.org/10.1371/journal.pone.0039191> PMID: 22723961
26. Belheouane M, Gupta Y, Künzel S, Ibrahim S, Baines JF. Improved detection of gene-microbe interactions in the mouse skin microbiota using high-resolution QTL mapping of 16S rRNA transcripts. *Microbiome*. 2017; 5: 59. <https://doi.org/10.1186/s40168-017-0275-5> PMID: 28587635
27. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, et al. Genetic Determinants of the Gut Microbiome in UK Twins. *Cell Host Microbe*. 2016; 19: 731–743. <https://doi.org/10.1016/j.chom.2016.04.017> PMID: 27173935
28. Blekhman R, Goodrich JK, Huang K, Sun Q, Bukowski R, Bell JT, et al. Host genetic variation impacts microbiome composition across human body sites. *Genome Biol*. 2015; 16: 191. <https://doi.org/10.1186/s13059-015-0759-1> PMID: 26374288
29. Bonder MJ, Kurilshikov A, Tigchelaar EF, Mujagic Z, Imhann F, Vila AV, et al. The effect of host genetics on the gut microbiome. *Nat Genet*. 2016; 48: 1407–1412. <https://doi.org/10.1038/ng.3663> PMID: 27694959
30. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*. 2011; 472: 57–63. <https://doi.org/10.1038/nature09922> PMID: 21475195
31. Herrema H, IJzerman RG, Nieuwdorp M. Emerging role of intestinal microbiota and microbial metabolites in metabolic control. *Diabetologia*. 2017; 60: 613–617. <https://doi.org/10.1007/s00125-016-4192-0> PMID: 28013341
32. Krautkramer KA, Kreznar JH, Romano KA, Vivas EI, Barrett-Wilt GA, Rabaglia ME, et al. Diet-Microbiota Interactions Mediate Global Epigenetic Programming in Multiple Host Tissues. *Mol Cell*. 2016; 64: 982–992. <https://doi.org/10.1016/j.molcel.2016.10.025> PMID: 27889451
33. Ridlon JM, Harris SC, Bhowmik S, Kang D-J, Hylemon PB. Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes*. 2016; 7: 22–39. <https://doi.org/10.1080/19490976.2015.1127483> PMID: 26939849
34. Romano KA, Martinez-Del Campo A, Kasahara K, Chittim CL, Vivas EI, Amador-Nogues D, et al. Metabolic, Epigenetic, and Transgenerational Effects of Gut Bacterial Choline Consumption. *Cell Host Microbe*. 2017; 22: 279–290.e7. <https://doi.org/10.1016/j.chom.2017.07.021> PMID: 28844887
35. Ridlon JM, Kang D-J, Hylemon PB. Bile salt biotransformations by human intestinal bacteria. *J Lipid Res*. 2006; 47: 241–259. <https://doi.org/10.1194/jlr.R500013-JLR200> PMID: 16299351
36. Wahlström A, Sayin SI, Marschall H-U, Bäckhed F. Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism. *Cell Metab*. 2016; 24: 41–50. <https://doi.org/10.1016/j.cmet.2016.05.005> PMID: 27320064
37. Kuipers F, Bloks VW, Groen AK. Beyond intestinal soap—bile acids in metabolic control. *Nat Rev Endocrinol*. 2014; 10: 488–498. <https://doi.org/10.1038/nrendo.2014.60> PMID: 24821328
38. Islam KBMS, Fukiya S, Hagio M, Fujii N, Ishizuka S, Ooka T, et al. Bile acid is a host factor that regulates the composition of the cecal microbiota in rats. *Gastroenterology*. 2011; 141: 1773–1781. <https://doi.org/10.1053/j.gastro.2011.07.046> PMID: 21839040
39. Zheng X, Huang F, Zhao A, Lei S, Zhang Y, Xie G, et al. Bile acid is a significant host factor shaping the gut microbiome of diet-induced obese mice. *BMC Biol*. 2017; 15: 120. <https://doi.org/10.1186/s12915-017-0462-7> PMID: 29241453

40. Begley M, Gahan CGM, Hill C. The interaction between bacteria and bile. *FEMS Microbiol Rev.* 2005; 29: 625–651. <https://doi.org/10.1016/j.femsre.2004.09.003> PMID: 16102595
41. Sayin SI, Wahlström A, Felin J, Jäntti S, Marschall H-U, Bamberg K, et al. Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. *Cell Metab.* 2013; 17: 225–235. <https://doi.org/10.1016/j.cmet.2013.01.003> PMID: 23395169
42. Svenson KL, Gatti DM, Valdar W, Welsh CE, Cheng R, Chesler EJ, et al. High-resolution genetic mapping using the Mouse Diversity outbred population. *Genetics.* 2012; 190: 437–447. <https://doi.org/10.1534/genetics.111.132597> PMID: 22345611
43. Churchill GA, Gatti DM, Munger SC, Svenson KL. The Diversity Outbred mouse population. *Mamm Genome.* 2012; 23: 713–718. <https://doi.org/10.1007/s00335-012-9414-2> PMID: 22892839
44. Kovacs A, Ben-Jacob N, Tayem H, Halperin E, Iraqi FA, Gophna U. Genotype is a stronger determinant than sex of the mouse gut microbiota. *Microb Ecol.* 2011; 61: 423–428. <https://doi.org/10.1007/s00248-010-9787-2> PMID: 21181142
45. O'Connor A, Quizon PM, Albright JE, Lin FT, Bennett BJ. Responsiveness of cardiometabolic-related microbiota to diet is influenced by host genetics. *Mamm Genome.* 2014; 25: 583–599. <https://doi.org/10.1007/s00335-014-9540-0> PMID: 25159725
46. Sehayek E, Hagey LR, Fung Y-Y, Duncan EM, Yu HJ, Eggertsen G, et al. Two loci on chromosome 9 control bile acid composition: evidence that a strong candidate gene, *Cyp8b1*, is not the culprit. *J Lipid Res.* 2006; 47: 2020–2027. <https://doi.org/10.1194/jlr.M600176-JLR200> PMID: 16763287
47. Keller MP, Gatti DM, Schueler KL, Rabaglia ME, Stapleton DS, Simecek P, et al. Genetic Drivers of Pancreatic Islet Function. *Genetics.* 2018; 209: 335–356. <https://doi.org/10.1534/genetics.118.300864> PMID: 29567659
48. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods.* 2016; 13: 581–583. <https://doi.org/10.1038/nmeth.3869> PMID: 27214047
49. Broman KW, Gatti DM, Simecek P, Furlotte NA, Prins P, Sen S, et al. R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multi-parent Populations. *Genetics.* 2018; genetics.301595.2018. <https://doi.org/10.1534/genetics.118.301595> PMID: 30591514
50. Org E, Parks BW, Joo JWJ, Emert B, Schwartzman W, Kang EY, et al. Genetic and environmental control of host-gut microbiota interactions. *Genome Res.* 2015; 25: 1558–1569. <https://doi.org/10.1101/gr.194118.115> PMID: 26260972
51. Davenport ER, Cusanovich DA, Michelini K, Barreiro LB, Ober C, Gilad Y. Genome-Wide Association Studies of the Human Gut Microbiota. White BA, editor. *PLoS One.* 2015; 10: e0140301. <https://doi.org/10.1371/journal.pone.0140301> PMID: 26528553
52. de Aguiar Vallim TQ, Tarling EJ, Edwards PA. Pleiotropic roles of bile acids in metabolism. *Cell Metab.* 2013; 17: 657–669. <https://doi.org/10.1016/j.cmet.2013.03.013> PMID: 23602448
53. Russell DW. The enzymes, regulation, and genetics of bile acid synthesis. *Annu Rev Biochem.* 2003; 72: 137–174. <https://doi.org/10.1146/annurev.biochem.72.121801.161712> PMID: 12543708
54. Martinot E, Sèdes L, Baptissart M, Lobaccaro J-M, Caira F, Beaudoin C, et al. Bile acids and their receptors. *Mol Aspects Med.* 2017; 56: 2–9. <https://doi.org/10.1016/j.mam.2017.01.006> PMID: 28153453
55. Inoue K, Zhuang L, Maddox DM, Smith SB, Ganapathy V. Structure, function, and expression pattern of a novel sodium-coupled citrate transporter (NaCT) cloned from mammalian brain. *J Biol Chem.* 2002; 277: 39469–39476. <https://doi.org/10.1074/jbc.M207072200> PMID: 12177002
56. Pesta DH, Perry RJ, Guebre-Egziabher F, Zhang D, Jurczak M, Fischer-Rosinsky A, et al. Prevention of diet-induced hepatic steatosis and hepatic insulin resistance by second generation antisense oligonucleotides targeted to the longevity gene *mIndy* (*Slc13a5*). *Aging (Albany NY).* 2015; 7: 1086–1093. <https://doi.org/10.18632/aging.100854> PMID: 26647160
57. Birkenfeld AL, Lee H-Y, Guebre-Egziabher F, Alves TC, Jurczak MJ, Jornayvaz FR, et al. Deletion of the mammalian *INDY* homolog mimics aspects of dietary restriction and protects against adiposity and insulin resistance in mice. *Cell Metab.* 2011; 14: 184–195. <https://doi.org/10.1016/j.cmet.2011.06.009> PMID: 21803289
58. von Loeffelholz C, Lieske S, Neuschäfer-Rube F, Willmes DM, Raschzok N, Sauer IM, et al. The human longevity gene homolog *INDY* and interleukin-6 interact in hepatic lipid metabolism. *Hepatology.* 2017; 66: 616–630. <https://doi.org/10.1002/hep.29089> PMID: 28133767
59. Li L, Li H, Garzel B, Yang H, Sueyoshi T, Li Q, et al. *SLC13A5* is a novel transcriptional target of the pregnane X receptor and sensitizes drug-induced steatosis in human liver. *Mol Pharmacol.* 2015; 87: 674–682. <https://doi.org/10.1124/mol.114.097287> PMID: 25628225

60. Staudinger JL, Goodwin B, Jones SA, Hawkins-Brown D, MacKenzie KI, LaTour A, et al. The nuclear receptor PXR is a lithocholic acid sensor that protects against liver toxicity. *Proc Natl Acad Sci U S A*. 2001; 98: 3369–3374. <https://doi.org/10.1073/pnas.051551698> PMID: 11248085
61. Civelek M, Lusis AJ. Systems genetics approaches to understand complex traits. *Nat Rev Genet*. 2014; 15: 34–48. <https://doi.org/10.1038/nrg3575> PMID: 24296534
62. Theriot CM, Bowman AA, Young VB. Antibiotic-Induced Alterations of the Gut Microbiota Alter Secondary Bile Acid Production and Allow for *Clostridium difficile* Spore Germination and Outgrowth in the Large Intestine. Ellerbe CD, editor. *mSphere*. 2016; 1: e00045–15.
63. Dawson PA, Haywood J, Craddock AL, Wilson M, Tietjen M, Kluckman K, et al. Targeted deletion of the ileal bile acid transporter eliminates enterohepatic cycling of bile acids in mice. *J Biol Chem*. 2003; 278: 33920–33927. <https://doi.org/10.1074/jbc.M306370200> PMID: 12819193
64. Oelkers P, Kirby LC, Heubi JE, Dawson PA. Primary bile acid malabsorption caused by mutations in the ileal sodium-dependent bile acid transporter gene (SLC10A2). *J Clin Invest*. 1997; 99: 1880–1887. <https://doi.org/10.1172/JCI119355> PMID: 9109432
65. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012; 337: 1190–1195. <https://doi.org/10.1126/science.1222794> PMID: 22955828
66. Chen J, Tian W. Explaining the disease phenotype of intergenic SNP through predicted long range regulation. *Nucleic Acids Res*. 2016; 44: 8641–8654. <https://doi.org/10.1093/nar/gkw519> PMID: 27280978
67. Janssen AWF, Dijk W, Boekhorst J, Kuipers F, Groen AK, Lukovac S, et al. ANGPTL4 promotes bile acid absorption during taurocholic acid supplementation via a mechanism dependent on the gut microbiota. *Biochim Biophys Acta*. 2017; 1862: 1056–1067.
68. Out C, Patankar J V, Doktorova M, Boesjes M, Bos T, de Boer S, et al. Gut microbiota inhibit Asbt-dependent intestinal bile acid reabsorption via Gata4. *J Hepatol*. 2015; 63: 697–704. <https://doi.org/10.1016/j.jhep.2015.04.030> PMID: 26022694
69. Miyata M, Yamakawa H, Hamatsu M, Kuribayashi H, Takamatsu Y, Yamazoe Y. Enterobacteria modulate intestinal bile acid transport and homeostasis through apical sodium-dependent bile acid transporter (SLC10A2) expression. *J Pharmacol Exp Ther*. 2011; 336: 188–196. <https://doi.org/10.1124/jpet.110.171736> PMID: 20884752
70. Boehm F. qtl2pleio: Hypothesis test of close linkage vs pleiotropy in multiparental populations. 2018.
71. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. *Annu Rev Psychol*. 2007; 58: 593–614. <https://doi.org/10.1146/annurev.psych.58.110405.085542> PMID: 16968208
72. Onishi JC, Campbell S, Moreau M, Patel F, Brooks AI, Zhou YX, et al. Bacterial communities in the small intestine respond differently to those in the caecum and colon in mice fed low- and high-fat diets. *Microbiology*. 2017; 163: 1189–1197. <https://doi.org/10.1099/mic.0.000496> PMID: 28742010
73. Li D, Chen H, Mao B, Yang Q, Zhao J, Gu Z, et al. Microbial Biogeography and Core Microbiota of the Rat Digestive Tract. *Sci Rep*. 2017; 8: 45840. <https://doi.org/10.1038/srep45840> PMID: 28374781
74. Northfield TC, McColl I. Postprandial concentrations of free and conjugated bile acids down the length of the normal human small intestine. *Gut*. 1973; 14: 513–518. <https://doi.org/10.1136/gut.14.7.513> PMID: 4729918
75. Hamilton JP, Xie G, Raufman J-P, Hogan S, Griffin TL, Packard CA, et al. Human cecal bile acids: concentration and spectrum. *Am J Physiol Gastrointest Liver Physiol*. 2007; 293: G256–63. <https://doi.org/10.1152/ajpgi.00027.2007> PMID: 17412828
76. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol*. 2013; 79: 5112–5120. <https://doi.org/10.1128/AEM.01043-13> PMID: 23793624
77. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010; 7: 335–336. <https://doi.org/10.1038/nmeth.f.303> PMID: 20383131
78. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013; 30: 772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
79. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. Poon AFY, editor. *PLoS One*. 2010; 5: e9490. <https://doi.org/10.1371/journal.pone.0009490> PMID: 20224823
80. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*. 2018; 6: 90. <https://doi.org/10.1186/s40168-018-0470-z> PMID: 29773078

81. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved GreenGenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 2012; 6: 610–618. <https://doi.org/10.1038/ismej.2011.139> PMID: 22134646
82. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol.* 2005; 71: 8228–8235. <https://doi.org/10.1128/AEM.71.12.8228-8235.2005> PMID: 16332807
83. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. Watson M, editor. *PLoS One.* 2013; 8: e61217. <https://doi.org/10.1371/journal.pone.0061217> PMID: 23630581
84. Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. *Nat Methods.* 2013; 10: 1200–1202. <https://doi.org/10.1038/nmeth.2658> PMID: 24076764
85. Scherer M, Gnewuch C, Schmitz G, Liebisch G. Rapid quantification of bile acids and their conjugates in serum by liquid chromatography-tandem mass spectrometry. *J Chromatogr B.* 2009; 877: 3920–3925. <https://doi.org/10.1016/j.jchromb.2009.09.038> PMID: 19819765
86. Morgan AP, Fu C-P, Kao C-Y, Welsh CE, Didion JP, Yadgary L, et al. The Mouse Universal Genotyping Array: From Substrains to Subspecies. G3 (Bethesda). 2015; 6: 263–279. <https://doi.org/10.1534/g3.115.022087> PMID: 26684931
87. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet.* 2014; 46: 100–6. <https://doi.org/10.1038/ng.2876> PMID: 24473328
88. Li Y, Tesson BM, Churchill GA, Jansen RC. Critical reasoning on causal inference in genome-wide linkage and association studies. *Trends Genet.* 2010; 26: 493–498. <https://doi.org/10.1016/j.tig.2010.09.002> PMID: 20951462
89. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol.* 1986; 51: 1173–1182. PMID: 3806354
90. Neto EC, Broman AT, Keller MP, Attie AD, Zhang B, Zhu J, et al. Modeling causality for pairs of phenotypes in system genetics. *Genetics.* 2013; 193: 1003–1013. <https://doi.org/10.1534/genetics.112.147124> PMID: 23288936
91. Rao A, Kusters A, Mells JE, Zhang W, Setchell KDR, Amanso AM, et al. Inhibition of ileal bile acid uptake protects against nonalcoholic fatty liver disease in high-fat diet-fed mice. *Sci Transl Med.* 2016; 8: 357ra122–357ra122. <https://doi.org/10.1126/scitranslmed.aaf4823> PMID: 27655848
92. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>( $\Delta\Delta C_T$ ) Method. *Methods.* 2001; 25: 402–408. <https://doi.org/10.1006/meth.2001.1262> PMID: 11846609
93. Clasquin MF, Melamud E, Rabinowitz JD. LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. *Curr Protoc Bioinforma.* 2012; Chapter 14: Unit14.11–14.11.23.
94. R Core Team. R: A Language and environment for statistical computing. R Foundation for Statistical Computing;
95. Harrell Jr FE, others with contributions from CD and many. *Hmisc: Harrell Miscellaneous.* 2018.
96. Kolde R. *heatmap: Pretty Heatmaps.* 2018.
97. McArdle BH, Anderson MJ. Fitting Multivariate Models to Community Data: A Comment on Distance-Based Redundancy Analysis. *Ecology.* 2001; 82: 290–297.
98. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlenn D, et al. *vegan: Community Ecology Package.* 2018.